Geoderma xxx (2016) xxx–xxx

Contents lists available at ScienceDirect

Geoderma

journal homepage: www.elsevier.com/locate/geoderma

Spatial disaggregation of complex Soil Map Units at the regional scale based on soil-landscape relationships

Sébastien Vincent, Blandine Lemercier *, Lionel Berthier, Christian Walter

UMR SAS, INRA, AGROCAMPUS OUEST, 35000, Rennes, France

ARTICLE INFO

Article history:
Received 5 January 2016
Received in revised form 1 June 2016
Accepted 4 June 2016
Available online xxxx

Keywords:
Spatial disaggregation
Digital soil mapping
Soil map units
Soil type units
Soil-landscape ruleset
Regional scale
Classification trees

ABSTRACT

Digital soil mapping is becoming a powerful tool to increase the spatial detail of soil information over large areas, which is essential to address agronomic and environmental issues. When it exists, information about soil is often sparse or available at a coarser resolution than required. The spatial distribution of soil at the regional scale is usually represented as a set of polygons defining Soil Map Units (SMUs), each including several Soil Type Units (STUs), which are not spatially delineated but semantically described in a database. Delineation of STUs within SMUs, i.e. spatial disaggregation of SMU, should improve the precision of soil information derived from legacy and ancillary data. The aim of this study was to predict STUs by spatially disaggregating SMUs through a decision-tree approach that considered expert knowledge about soil-landscape relationships embedded in soil databases. In a 27,376 km^2 study area in north-western France (Brittany), 434 SMUs were delineated at 1:250,000 scale, and 320 STUs, their relative area in the SMUs, and their geomorphological and geological contexts were described. A calibration dataset of points was established using stratified random sampling (n = 352,188). To retrieve soil-landscape relationships, expert rules for soil distribution defined by soil surveyors and based on topography, parent material and waterlogging index were considered in order to allocate an STU to 83% of the calibration dataset. The calibration dataset and covariates (i.e. pedological, geological and terrain attributes; land use; airborne gamma-ray spectrometry) were then used to build and extrapolate the decision tree using the C.5 algorithm in DSMART software. Several iterations were performed, providing a probability of occurrence of each possible STU within the study area. External validation was performed by comparing predictions of the disaggregation procedure to available soil maps at scales of 1:25,000 or 1:50,000 and observed profiles. Overall accuracies ranged from 41 to 72%, depending on the validation method (per pixel vs. 3 × 3 windows of pixels, per STU vs. STU grouped by semantic proximity (n = 204)). Introducing expert rules based on soil-landscape relationships to allocate STUs to calibration samples enabled production of a soil map with clear spatial structures, yielding expected spatial patterns of soil organisation. Future work notably concerns estimating soil properties at multiple depths deriving from STU predictions, according to the GlobalSoilMap project.

1. Introduction

Accurate and precise soil information over large areas is essential to manage agronomic and environmental issues and help decision makers in regional management (Brungard et al., 2015). When it is available, information on soil is often sparse or available at a coarser resolution than required. At regional or country levels, soils are usually mapped at an exploratory scale (1:250,000–1:500,000) (Hartemink, 2008). Variability in soil types is too great to be mapped at this scale. Soils are thus represented as a set of polygons with sharp boundaries - Soil Map Units (SMUs) - each containing several Soil Type Units (STUs), which are not spatially delineated but associated with a semantic database describing them, with various degrees of precision according to analysis date, author and/or national and local specifications. It is an aggregation technique which requires descriptive information about the spatial organisation of STUs within SMUs and full extent landscape descriptors. Thus, to keep a spatial representation at 1:250,000 scale consistent, STU information is aggregated at the SMU level.

Soil maps at 1:250,000 provide an overview of the soil of a large area and are a compromise between a detailed and an exploratory soil map in terms of cost and accuracy. This aggregated information is insufficient, however, for many uses, as the delineation of wetlands or the valuation of ecosystem services rendered by soils for instance. More precise representation and a change to a geographical support that allows the use of qualitative and quantitative data at the elementary-unit level (STU) is needed.

In recent decades, Digital Soil Mapping (DSM) has become the main operational tool to produce soil data over large areas and at a resolution usually finer than used in legacy polygon mapping (Lagacherie et al.,

* Corresponding author at: Agrocampus Ouest, UMR 1069 SAS, F-35042 Rennes, France.
E-mail address: blandine.lemercier@agrocampus-ouest.fr (B. Lemercier).

http://dx.doi.org/10.1016/j.geoderma.2016.06.006
0016-7061/© 2016 Published by Elsevier B.V.

2013). The rise of DSM techniques is explained mainly by the large amounts of time and resources needed for traditional soil mapping, as well as by the development of Geographic Information Systems (GIS) and statistical methods, and the availability of spatially-explicit data about the environment (Grunwald, 2009; McBratney et al., 2003). A high-resolution grid of soils and their properties is the goal of the GlobalSoilMap project (Arrouays et al., 2014; Sanchez et al., 2009). The GlobalSoilMap consortium aims to make available to end-users (e.g., farmers' associations, policy makers, research communities, and agribusinesses) a database of soils and their properties at several depths, and uncertainties associated with estimated soil properties. The GlobalSoilMap products are suitable for spatial and temporal modelling needing information about stores and fluxes of water, carbon, nutrients and solutes in soils (Arrouays et al., 2014).

There are many approaches to produce soil data in raster format using DSM techniques. In a review, Grunwald (2009) analysed 90 DSM publications from 2007 to 2008 from two international soil journals: several methods were used in different areas in the world. It appeared that each study is unique, and one single approach cannot address all situations. DSM methods are currently used to downscale soil information. Two main pathways can be explored to increase the precision or resolution of maps: applying geostatistical models using point input data or spatial disaggregation of complex SMUs. Map information is composed of mapping units containing one or more soil components that can be mapped separately. Spatial disaggregation can be defined as the process of separating an entity into component parts based on implicit spatial relationships or patterns. Techniques to spatially disaggregate information include assumptions about the spatial distribution of the target variable or the relationship between the target variable and the auxiliary geographical variables (Vogel et al., 2015).

Spatial disaggregation is commonly used in climatic and hydrological studies to move from a global model to regional and local impact models (Boe et al., 2007; Skaugen, 2002). In pedology, the concept of spatial disaggregation was introduced in 1998 (McBratney, 1998) and has been developed in recent years in many studies around the world, primarily to move from polygon-based maps to raster-based format, compatible with current needs (Nauman and Thompson, 2014). Several methods have been developed, depending on data format and availability.

Tree-based models are increasingly used in digital soil mapping studies (Bui and Moran, 2001; Lacoste et al., 2011; Haering et al., 2012; Lemercier et al., 2012; Heung et al., 2014; Nauman and Thompson, 2014; Odgers et al., 2014b; Vaysse and Lagacherie, 2015) due to their ability to transcribe the complex relationship between soil and landscape (Walter et al., 2006), their ease of use (Grunwald, 2009), and their ability to handle qualitative and quantitative environmental variables (Scull et al., 2005). Other authors have built decision trees to predict soil units or soil groups in regions such as a French valley of the Languedoc-Roussillon region (Lagacherie and Holmes, 1997), Denmark (Adhikari and Hartemink, 2016), and the Mojave Desert in California, United States (Scull et al., 2005). Decision-tree approaches were also used to map geological classes (Bui and Moran, 2001). Odgers et al. (2014b) applied a classification decision tree with covariates to disaggregate a soil polygon map in central Queensland, Australia. A stratified random sampling dataset was designed to calibrate the decision tree, and soil types were allocated to each sample using weights. The same procedure was applied by Holmes et al. (2015) to predict soil groups in western Australia. They concluded that one way to significantly improve prediction accuracy could be to add rules about the relationship between soil and landscape into the sampling procedure.

Nauman and Thompson (2014) introduced expert rules to spatially disaggregate SMUs, using soil information retrieved from a database and environmental covariates. In their study, each landform situation was described and translated into raster rules. In another approach, Haering et al. (2012) used soil-landscape relationships with topographic attributes to separate soil types within each complex SMU; for each soil type, a topographic "fingerprint" was made by analysing the distribution of topographical features among the soil types.

Our article describes recent work on spatial disaggregation of existing SMUs covering the Brittany region in north-western France to predict and map the STUs in a raster-grid format at a resolution of 50 × 50 m. Expert rules based on soil-landscape relationships were retrieved from the existing soil database and then integrated into the prediction procedure and combined with covariates in a classification tree scheme.

2. Materials

2.1. Study area

The study area is the Brittany region in north-western France (Fig. 1). Its 27,040 km^2 have diverse physical and geographic conditions. Its climate is oceanic, with mild temperatures that do not vary greatly throughout the year, with mean annual precipitation of 500–1500 mm and mean annual temperature of 10.5–13 °C. Brittany's climate is not homogeneous, however, having high variability between coastal and inland zones. Consequently, agricultural land use varies, with vegetable production near the Northern coasts and crop-livestock production inland.

The relief is gentle and highly correlated with geological formations; elevation varies from 0 to 400 m. Brittany is part of the Armorican Massif, whose geology is complex: basement formation (granite, gneiss and micaschist) in northern and southern zones, sedimentary rocks (sandstone, Brioverian schist) in the central zone, and superficial deposits (Aeolian loam, alluvial and colluvial deposits) overlaying bedrock formations. This high geological diversity generates a rich mosaic of soils. The main soils are Cambisols, Stagnic Fluvisols and Leptosols. Soils are mainly organised along toposequences: the slopes, generally well drained, carry soil varying in thickness from top to bottom of the slope; in valleys, redoximorphic soil are common.

2.2. Soil data

i) Regional 1:250,000 soil database

A comprehensive soil map and information about soil spatial organisation are required to implement sustainable soil management in the region. In France, regional geographic databases at 1:250,000 scale, called "Référentiels Régionaux Pédologiques (RRP)" have been developed according to national specifications (INRA Infosol, 2005). Point and surface data of soil studies are organised and gathered in a national database: DoneSol (INRA Infosol, 2014). In Brittany the initiative was applied through the "Sols de Bretagne" programme, started in 2005 and certified in 2012. Soil is represented as a set of polygons defining SMUs (Fig. 1), each one including 1 to several STUs, not spatially delineated, and a semantic database describing them. An average of 5.8 polygons is assigned to each SMU. SMUs are the smallest pedological entities representable at the target scale. They are defined as a part of the landscape in which soil genesis factors are homogeneous (e.g., morphology, geology, climate, and, in some cases, land use). An STU is a portion of the soil cover which has identical pedogenesis and, at any point in space, the same sequence of diagnostic horizons. In the study area, 341 SMUs and 320 STUs have been identified (mean = 7 STU per SMU), from a total of 1984 polygons, each with a mean area of 1.36 km^2.

STU boundaries cannot be represented at 1:250,000, but the organisation of soil types in each SMU was described as accurately as possible in the DoneSol database and in documents on soil organisation published for each department of Brittany (Berthier et al., 2013a,b; Le Bris et al., 2013a,b).

Please cite this article as: Vincent, S., et al., Spatial disaggregation of complex Soil Map Units at the regional scale based on soil-landscape relationships, Geoderma (2016), http://dx.doi.org/10.1016/j.geoderma.2016.06.006